The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #42813

Awareness of Feature Importance in Artificial Intelligence Algorithms

Dr. Ebisa Wollega, Colorado State University, Pueblo

Ebisa Wollega, Ph.D., is an Associate Professor of Engineering at Colorado State University Pueblo. His research interests include applied artificial intelligence, large-scale optimization, and engineering education.

Melissa Braddock
Dr. Lisa Bosman, Purdue University, West Lafayette

Dr. Bosman holds a PhD in Industrial Engineering. Her engineering education research interests include entrepreneurially minded learning, energy education, interdisciplinary education, and faculty professional development.

Awareness of Feature Importance in Artificial Intelligence Algorithms

Abstract: Industrial engineering graduates need to be familiar with artificial intelligence (AI) due to its transformative impact on modern manufacturing and production processes. AI technologies, such as machine learning and predictive analytics, optimize resource allocation, enhance efficiency, and streamline operations. Proficiency in AI equips graduates to innovate, automate tasks, and address complex industrial challenges effectively. Predictive models are typically taught in one or more Industrial Engineering courses, such as Operations Planning and Control at Colorado State University Pueblo. It is beneficial that students learn the general implications of using predictive models. The models utilize various AI algorithms, where an algorithm learns from a retrospective dataset that comprises samples and features to make predictions. An AI algorithm trained on a balanced dataset produces good results, while an algorithm trained on a biased dataset may lead to unfavorable outcomes. Some dataset features can be eliminated as insignificant, and the AI algorithm is deployed only on the significant features. There are many reasons why features are insignificant beyond using just biased datasets. However, it is worthwhile to investigate the effects of these insignificant features, as detailed analyses can reveal damaging or positive consequences of the omitted features. In this paper, three publicly accessible datasets are used to present subjective analyses of insignificant features beyond the general recommendation of an AI algorithm.

Introduction

Predictive modeling is one of the skills that industrial engineering graduates possess. The predictive relevance of an output/response depends on the quality of the input factors, commonly known as the features. Among many input factors, choosing significant features for business operational purposes is customary. There are many reasons why a portion of the features is used for decision-making purposes, such as computational cost and accuracy [1,2,3]. In addition, the predictive algorithms are also tuned carefully to optimize for overfitting and underfitting [4]. The selection of significant features is based on established theories that are well-accepted in practice [1,2,3]. However, in this paper, we use the widely used p-value technique in statistics to determine the significance of a factor. Undergraduate students relatively easily understand a linear regression model and the p-value technique compared to advanced machine learning feature selection algorithms [1,2,3] presented in the literature.

We perform significance analyses of three datasets used in the Operations Planning and Control course at Colorado State University Pueblo University and attempt to tell the stories of the insignificant features. These are the Advertising, Auto, and Credit Card Balance datasets, and a linear regression predictive model is commonly used to analyze the datasets. Telling the stories will expose students to a panorama of subjective views of the data analysis results to make inclusive and robust decisions. There are ongoing attempts to look directly or indirectly into the insignificant factors in the form of AI ethics [5,6,7] and AI fairness [8,9]. For example, AIFairness360 (AIF360) software is available to detect, understand, and mitigate algorithmic bias [10]. While we highly encourage educators to explore the capability of AIF360, we use the datasets without manipulation to compute the p-values of the features.

We want the reader to know that the purpose of this paper is neither to present descriptive nor predictive analyses of the datasets used. Interested readers can refer to the datasets' sources in the reference section [11], where the p-values are computed, the regression equations are presented, and the data summary is presented. We aim to select one insignificant feature from each dataset to show-case students how to go beyond what the p-values tell us and perform qualitative analysis instead of just using what our AI algorithms tell us about the dataset.

In the following sections, we first briefly present each dataset's purpose, sample size, and input features. Then, we select an insignificant feature generated by an AI algorithm when a multiple linear regression model is deployed. Lastly, we present a qualitative analysis of an insignificant feature. We do not provide detailed explanations for the input features and the regression equation, nor cover the preprocessing of the data. Again, readers can consult the reference materials [11] for the detailed computational aspects. ***Auto*** and ***Credit Card Balance*** datasets are part of the ISLR package, and the regression model can be run in the RStudio. The ***advertising*** dataset is a solved example in the reference material [11]. We have carefully chosen the data from publicly available libraries and textbooks [11] so that educators can easily access the datasets to use in their classrooms. For readability purposes, we write a ***dataset*** in ***italic bold*** and the *features* of the dataset just in *italic* fonts.

Advertising Dataset

The purpose of the ***advertising*** dataset is to estimate the sales of a product in thousands of units for a business franchise when thousands of dollars are spent on *TV, radio, and newspaper* advertising. The linear regression analysis results are presented in the reference book [11]. We call the sales volume the response and the *TV, radio, and newspaper* advertising media features. The dataset has two hundred samples, which represent two hundred different markets. A multiple linear regression predictive AI model is used to identify the significance of the features.

The p-value analysis indicates the *newspaper* is an insignificant feature, implying that the *newspaper* feature shall not be used to predict the sales volume. As a result, the franchise will not pay money to advertise their product with the newspaper publishers since the AI model suggests that the impact of the *newspaper* feature is significantly minimal. From a business decision point of view, not considering the newspaper as an advertisement channel makes sense since the significant features highly contribute to improving the sales volume. It also implies that the franchise would pay for the TV and radio to run the advertising, which benefits both broadcasting services. However, the franchise would halt paying the newspaper for advertising, which could lead to the demise of the newspaper printing services.

The newspaper publishing discontinuity is the story behind the insignificant features that we want our students to be aware of when using AI output for decision-making. The story needs further subjective analysis beyond the predictive quantitative results of the AI recommending significant features to make business decisions. For example, the publishing enterprise employee may have been in the publishing business for decades, and it could be too late to switch jobs to maintain their livelihood, resulting in increased poverty and dependency on government subsidies. In addition, the closure of publishing enterprises could lead to a lack of access to newspapers, which could lead to a reduced quality of life for those who grew up reading non-online publications.

Auto Dataset

This ***Auto*** dataset aligns with Quinlan 1993[12] to predict miles per gallon (mpg) for city-cycle fuel consumption based on eight input features: engine *displacement* in cubic inches, number of *cylinders* between 4 and 8, engine *horsepower*, vehicle *weight* in lbs., time of *acceleration* in seconds from 0 to 60 miles per hour, model *year*, whether the *origin* of car is American, European, or Japanese, and a vehicle *name* [11]. Three hundred ninety-two instances of the data are used for our analysis. The dataset was initially used in the 1983 American Statistical Association Exposition and maintained at Carnegie Mellon University for the StatLib library.

When feature significance analysis is performed using a multiple linear regression predictive model, the p-value shows that the *acceleration* feature is an insignificant factor, which means the feature will be excluded from the estimation of city-cycle fuel consumption of a vehicle. Excluding the acceleration feature may make sense from the AI model's application point of view. Auto manufacturers can also reallocate the resources used to collect acceleration data to something else for efficiency.

However, as educators, it is imperative that we encourage students to dive deep into understanding the pros and cons of the features overlooked by our AI algorithm.
For example, motor vehicle acceleration can impact public safety, especially when vehicles share roads with pedestrians, bicyclists, and animals. According to the National Highway Traffic Safety Administration, a pedal application error mistaking for the brake has led to fifteen crashes per month in the United States [13]. Modern cars can reach 60 miles per hour from 0 miles per hour in less than 3 seconds, and it is also customary that auto manufacturers desire maximum acceleration as a selling point. For example, Bloomberg reported the following in 2023, under the title Extreme Acceleration Is the New Traffic Safety Frontier, based on an interview with an expert. "*Auto companies have long insisted that strong acceleration can be a safety feature, as when passing a car or merging onto a freeway…that can be true….*" In addition, they also reported that "*…cars' huge jump in launch capability poses a potential threat in an urban environment, where vehicles in close quarters must share space with pedestrians, bicyclists, and other road users.*"

The story's moral is that acceleration's general relevance cannot be overlooked, even though our predictive model tells us otherwise for the miles-per-gallon estimation. In fact, there are Acceleration Safety Mode technologies such as the "Eco," "Normal," and "Sport" modes being integrated into modern cars, where the Eco mode reduces throttle reaction time to save fuel.

Credit Card Balance Dataset

The ***credit card balance*** dataset is simulated data containing information for four hundred customers [11]. The purpose of the dataset is to predict a customer's credit card balance in dollars to indicate which customers will default on their credit card debt. The dataset comprises samples with the following features: the customer's *income* in thousands of dollars, the credit *limit* of the customer, the credit *rating* of the customer, the number of credit *cards* the customer owns, *age* in years of the customer, number of years of *education*, *gender* with labels male and female, currently if a customer is a *student* or not, whether a customer is *married* or not, and if *ethnicity* is African American, Asian, or Caucasian. The predicted outcome can be used to make critical decisions, such as mortgage loan approvals.

The p-value analysis shows that *gender* is an insignificant feature for a multiple linear regression predictive model, which means that the gender input will not be used directly to predict a customer's credit card balance. You can refer to an analysis of the dataset on Kaggle, as an example. This can be a great relief for the data analyst since gender is a protected status information, and fairness through blindness [14] appears to be achieved. However, students have to be encouraged to explore other views, such as fairness through awareness [15,16,17], since some significant factors, such as income, can correlate with gender. This may imply that our AI algorithm is not necessarily "blind" to the protected status of gender. For example, a person on maternity/paternity leave can be unemployed during the period, resulting in a reduced *income*; however, they can be a high earner when they return to work. For example, Forbes reports on a similar issue in 2022 titled, "Is It Time To Start Using Race And Gender To Combat Bias In Lending?" A few lines read, "*The assumption behind Equal Credit Opportunity Act passed in 1974 was that if decision makers-be they humans or machines-are unaware of attributes like race or gender at decision-time, then the actions they take will be based on 'neutral' and 'objective' factors that are fair. There's just one problem with this assumption: It's wishful thinking to assume that keeping algorithms blind to protected characteristics means the algorithms won't discriminate. In fact, building models that are 'blind' to protected status information may reinforce pre-existing biases in the data.*"

The AI algorithm could be trained to normalize the analysis time window for possible maternity/paternity leave. There is ongoing research to mitigate bias to improve fairness in AI. For example, the AI Fairness 360 machine learning Python toolkit is designed to understand and mitigate possible bias in AI algorithm-based decision-making [10]. Techniques are also proposed to improve robustness to minimize bias in training the AI algorithms [18, 19].

Conclusions

In this paper, we make a case for not overlooking insignificant features when making decisions using AI algorithms. It is crucial not to overlook insignificant features when making decisions using AI algorithms because even seemingly minor variables can have an impact on the accuracy and reliability of the model's predictions. Neglecting these features could lead to biased or flawed outcomes, undermining the effectiveness and trustworthiness of the AI system. Additionally, including all relevant features ensures comprehensive analysis and minimizes the risk of overlooking potential patterns or correlations that could be valuable for decision-making.

As educators, this is particularly relevant to improving students' critical thinking and constantly reminding students to have a holistic picture of data analysis results before making critical decisions. In addition, we need to remind business owners to abide by their core values to serve their community when making critical decisions based on AI algorithms. We can accomplish this goal by educating our students, who are future engineers, leaders, business owners, and decision-makers in general. Overall, the content presented in this paper can be used as an example in course instruction to illustrate the relevance of subjective analysis in a work environment when qualitative analysis of data drives decision-making.

The datasets used in this paper were not altered or transformed to fit the multiple linear regression model better. However, the insignificant features can become significant if a larger dataset is used, the number of features is increased or decreased, data transformation is performed, and an AI model other than linear regression is used. However, to be aware of a balanced approach in courses that utilize AI algorithms is recommended.

References

[1] Kumar, V., & Minz, S. (2014). Feature selection: a literature review. *SmartCR*, *4*(3), 211-229.
[2] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, *50*(6), 1-45.
[3] Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249-256). Morgan Kaufmann.
[4] Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, *70*(10.3850), 978-981.
[5] Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management (JDM)*, *31*(2), 74-87.
[6] Hagerty, A., & Rubinov, I. (2019). Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.
[7] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.
[8] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, *54*(6), 1-35.
[9] Mahoney, T., Varshney, K., & Hind, M. (2020). *AI fairness*. O'Reilly Media, Incorporated.
[10] Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, *63*(4/5), 4-1.
[11] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) *An Introduction to Statistical Learning with applications in R*, https://www.statlearning.com, Springer-Verlag, New York
[12] Quinlan, R.. (1993). Auto MPG. UCI Machine Learning Repository. https://doi.org/10.24432/C5859H.
[13] Lococo, K. H., Staplin, L., Martell, C. A., & Sifrit, K. J. (March 2012). Pedal Application

Errors. (Report No. DOT HS 811 597). Washington, DC: National Highway Traffic Safety Administration.
[14] Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of behavioral decision making*, *8*(2), 71-83.
[15] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012, January). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (pp. 214-226).
[16] Zhang, W., Hernandez-Boussard, T., & Weiss, J. (2023, June). Censored fairness through awareness. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 12, pp. 14611-14619).
[17] Greenberg, J. (1983). Overcoming egocentric bias in perceived fairness through self-awareness. *Social Psychology Quarterly*, 152-156.
[18] Ghosh, S., Squillante, M., & Wollega, E. (2021). Efficient Generalization with Distributionally Robust Learning. *Advances in Neural Information Processing Systems*, *34*, 28310-28322.
[19] Kuhn, D., Esfahani, P. M., Nguyen, V. A., & Shafieezadeh-Abadeh, S. (2019). Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations research & management science in the age of analytics* (pp. 130-166). Informs.